Trade-offs Between Immersion and Energy Consumption With Real-Time Light Estimation in Augmented Reality: A Pre-Study

STEFANO FORMICOLA   CHRISTOPH ALBERT JOHNS

`formico` | `cajohns` `@kth.se`

January 15, 2021

Abstract

This research project investigates the effect of real-time light estimation in augmented reality (AR) applications on immersion and on energy consumption in a pre-study with seven participants. A simple open-source card-matching game is modified to implement and control the corresponding rendering setting as provided by Apple's AR application programming interface (API) RealityKit and tested with a small convenience sample of users. A within-subject research design is chosen to measure immersion through the Immersive Experience Questionnaire (IEQ) while overall central processing unit (CPU) usage is monitored and logged for both test conditions in randomized order. A Mann-Whitney-U-test is used to test for significant differences in immersion and CPU usage between both conditions. The preliminary results seem to suggest that immersion is not significantly affected by real-time light estimation whereas CPU usage might be higher for the enabled condition, although a larger scale study is required for a more meaningful account. For such a large-scale study, usability issues of the sample application should be resolved. Additionally, it should be considered whether the IEQ should be replaced with a questionnaire that is more specific to AR applications.

Keywords: Augmented reality, Lighting, Immersion, Energy consumption

Contents

List of Acronyms and Abbreviations

API application programming interface

AR augmented reality

CPU central processing unit

GPU graphics processing unit

IDE integrated development environment

IEQ Immersive Experience Questionnaire

1 Introduction

Through the dissemination of powerful and easy-to-use framework APIs, increasingly advanced features for AR have become readily available to developers and designers. Apple's AR API *RealityKit*, for example, has enabled creators to easily add occlusion, video textures, and shared states for collaborative experiences to their AR applications [1]. One of the especially notable new features of recent years has been real-time light estimation [2]. Applications that enable this feature—or, as in the case of RealityKit, choose to keep this default setting—gain access to information about a scene's lighting with every video frame that is delivered to the current session and can use this to apply dynamic shading to virtual objects matching the real-world lighting conditions [2]. To achieve this, camera-sensor data for one or multiple probes in the scene is continuously processed and evaluated [2, 3]. This, depending on the scene and lighting conditions, can be a computationally intensive operation [4]. Enabling real-time light estimation could, therefore, lead to increased scene realism at the cost of decreased energy efficiency.

This research aims at illuminating the relationship between immersion and energy consumption in deploying real-time light estimation, thus empowering developers and designers to decide the most appropriate feature sets for their applications considering both user experience and sustainability aspects. For this, our project builds on two main lines of research: prior works evaluating the relationship between visual characteristics of AR objects and user experience (e.g. [5]) and research on the approximation of natural lighting conditions within virtual scenes in augmented reality (e.g. [6]). While there has been extensive research on the effect of design decisions on several aspects of user experience in digital games [7]—even in AR games (e.g. [8])—the effect of lighting and more specifically real-time light estimation in AR games on immersion seems not to have been examined yet.

We chose to investigate the effect of real-time light estimation on immersion—in contrast to, for example, *flow* [9]—because it is independent from optimal experience, which our method wasn't able to produce, and because it is very likely to relate to visual presentation and, therefore, lighting [10]. There exist multiple alternative measures that could have been chosen to explore aspects of the experience when playing video games instead (see [11, 12] for overviews). We reason, however, that immersion is an appropriate measure for this study since real-time light estimation is likely to enhance the sense of realism of a scene. As Witmer and Singer argue, this sense of "scene realism" is one of the governing factors for presence in virtual environments [13], which in turn affects how immersive a virtual environment appears [10].

Since we were limited by our resources in terms of time and budget, however, and since this research project was conducted during the COVID-19 pandemic, we were not able to reach the necessary number of participants that would be required to fully illuminate said relationship. Instead, we focused on conducting a smaller scale pre-study that can give some indication about the results as well as provide some insights into the suitability of our chosen methods and tools. The deliverables of our project are, therefore: an example AR application to measure immersion levels and monitor energy consumption, a tested and evaluated method (setup, procedure, and analysis) for a full-scale study, and an indication for the results of said study.

1.1 Research Questions

For this pre-study as well as for the eventual full-scale study, we are interested in two major research questions relating to one of the two aspects of our investigation each: real-time light estimation's effect on immersion and its effect on energy consumption. More specifically, we want to examine users' perceived immersion when they are explicitly asked to rate said experience using a standardized questionnaire (i.e. the IEQ developed by Jennett et al. [10]) as well as their own unsupported articulation of any difference between enabled and disabled real-time light estimation. Additionally, we want to explore the effect of real-time light estimation on device metrics collected in the background that can give an indication about the condition's effect on device resource utilization. Overall, we proposed to investigate the following questions:

RQ1. Do users experience a difference in immersion between enabled and disabled real-time light estimation?

RQ1a. If they perceive a difference, are the participants able to tell that lighting is the cause?

RQ2. How does enabling real-time light estimation affect CPU usage?

In our pre-study, energy consumption has been measured in terms of CPU usage because it can very finely be measured on the test device in contrast to general energy consumption.

1.2 Research Hypotheses

While we are aware that we were only able to test the proposed method on a small scale, we hypothesize the following outcomes for a full-scale experiment:

H1. Users experience higher immersion levels when real-time light estimation is enabled.

H1a. Users are able to tell that the lighting was different between both runs of the experiment.

H2. CPU usage is higher when real-time light estimation is enabled.

As has been explained above, scene realism is positively correlated with presence and thereby immersion. That is why we expect the lighting condition that more closely matches the scene's natural environment to lead to higher levels of immersion, resulting in the first hypothesis. Second, because the difference between both conditions lies in a salient visual feature which affects AR objects' color and reflections, we expect not only the indirect measure of immersion as assessed through the IEQ to be affected. We instead hypothesize that users will recognize the difference between both conditions and will be able to articulate how they differ without support from the examiner. And finally, regarding the energy impact of real-time light estimation, we expect the computationally intensive continuous probing and estimation of the AR scenes environment lighting to lead to higher levels of overall CPU usage.

2 Method

We propose a within-subject design to test the effects of the two conditions *real-time light estimation enabled* and *real-time light estimation disabled* on immersion using the IEQ developed by Jennett et al. [10] while measuring CPU usage via a device metrics logging tool in combination with a sample AR application. For this, we modified an existing open source mobile AR game and prepared a form that includes the IEQ, while Xcode's Developer Tools' *Instruments* were used to measure CPU usage over time. Because of its popularity across countries and languages, we considered the game *Memory* or *Concentration*, in which players have to remember the order and color of different cards, a valid candidate for the experiment. The implementation was done for iOS devices and developed on Apple's integrated development environment (IDE) Xcode using the ARKit and RealityKit APIs. The experiment consisted of one three-minute session of gameplay for each condition, with and without the real-time light estimation, per subject. The user interface included a toggle button hidden from the player to enable and disable the real-time light estimation feature for the two runs of the experiment (see Figure 1 for a comparison between enabled and disabled real-time light estimation in the sample application). In the following, the participant profile, procedure, data collection, and analysis are described.
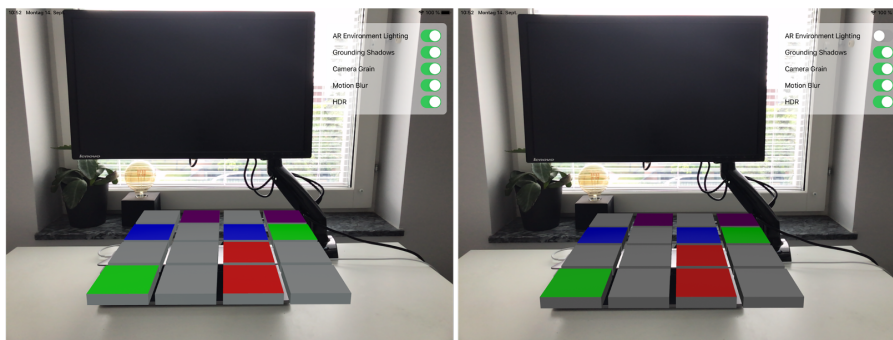

Figure 1: Comparison between enabled (left) and disabled (right) real-time light estimation in the sample card-matching AR application. The controls can be hidden before the test device is handed over to the participant.

2.1 Participants

The participants were recruited via an opportunity sample and were—due to the difficulties and restrictions regarding gatherings of people during the COVID-19 pandemic—mostly international students living on

campus at a major technical university in Sweden. To limit the effect of novelty on immersion levels, they were required to have had played a card-matching game similar to the sample game before. It was, however, not required for participants to have played a digital version of this game before taking part in the experiment. To further mitigate the effect of novelty, participants were required to have experienced an AR scene on a mobile handheld device before. Overall, seven participants took part in this study. They had an average age of 24.7 (SD = 1.60), ranging from 23 to 28 years. All seven participants were male.

2.2   Preparations

To be able to conduct the proposed experiment, we prepared an example application that could easily control the currently active lighting setting and a digital questionnaire that included a consent form as well as the IEQ.

We modified an open-source AR card-matching game [14] to include an option to enable and disable real-time light estimation using RealityKit's *.disableAREnvironmentLighting* render option. The application was then installed on an Apple iPad Pro 2017 (model number FPDY2FD/A) so it could be presented to participants and monitored using the Xcode IDE. The application that formed the basis for our experiment was chosen due to its simplicity in functionality and visual presentation as well as participants' general familiarity with the underlying game of *Concentration*, thereby reducing the risk of errors in use or unwanted effects on immersion due to scene complexity or novelty. Additionally, several suitable rooms accessible to students were identified so the experiment could be conducted in an indoor setting with controlled lighting.

To collect consent as well as immersion data using the IEQ, a Google Forms document was prepared that informed participants about the purpose of the experiment as well as the kind of data that would be collected, stored, and analyzed (see Appendix A). Additionally, it included two instances of the IEQ developed by Jennett et al. [10] to collect data on players' immersion levels for each of the two lighting conditions. The full extent of the IEQ as used in this experiment can be found in Appendix B.

2.3   Procedure

Before the start of the experiment, participants were informed about the collection of data regarding the device activity and their questionnaire answers after both runs of the game. Moreover, their right to withdraw from the experiment at any time was clarified as well as the possibility to ask for the deletion of all data associated with their person. The purpose of the experiment was stated to the participants in general terms as "investigating immersion in AR games" to avoid influencing their perception of the scene by directing their attention towards lighting. They were asked to fill the consent form detailed above.

To further mitigate the risk of influencing participants and to address the concerns brought up by Jennett et al. [10] regarding the effect of prior experience on immersion, participants were then given a trial period with a third lighting condition: disabled real-time light estimation and an added point light above the scene. After they had indicated to sufficiently understand the experiment application, participants were asked to complete as many rounds of the card-matching game with as few errors as possible, before they were interrupted after three minutes and the render option was changed for the second run following the study design of Jennett et al. [10]. The motivation for this experiment design was to allow for a more challenging experience across player ability levels, thereby increasing flow [9] and general immersion levels as well as reducing overall variance in immersion [10]. The order of both render options was randomized to improve reliability and participants were not informed about the currently active render option to avoid biases. The first run was considered started after the participants indicated that they were ready for the experiment to begin.

During each run of the experiment, the IDE Xcode and more specifically its *Instruments* feature [15] were used to log the test device's CPU utilization in three second intervals according to Apple's type definitions [16]. While there exists a more general power metric that can be logged through the same application [17], that metric is discrete, abstract, and too inaccurate to be elastic towards enabled and disabled real-time light estimation and was, therefore, not used. We decided not to have participants think aloud during their app usage to avoid unwanted effects on their immersion [18].

After each run, participants were asked to fill out the IEQ detailed above to measure their levels of immersion dependent on the lighting treatment. After the second run, participants were additionally asked about their general experience using the application and to elaborate on any differences they noticed between both runs. While alternatives such as tracking eye movement or measuring reaction times could also have been used to measure immersion [10], we chose a questionnaire approach as the most economical option that still has

satisfactory precedence in literature [19]. If a participant referred to "light", "lighting", "color", or "contrast" while answering the open questions in the questionnaire, the difference in lighting was considered as noticed by the user.

After the experiment was completed, participants were debriefed and fully informed about the purpose of the study and given the opportunity to withdraw their consent. The duration of the experiment per participant was around twenty minutes.

2.4   Data Analysis

After the above data had been successfully gathered, we conducted a statistical analysis regarding the participants' immersion levels and the CPU usage between both lighting conditions using a Mann-Whitney U-test to check for significant differences. We chose this test because it does not require a normal distribution and because it was used by Jennett et al. in their original studies on immersion [10]. However, the Student's t-test's could also have been used—as for example Nordin et al. [20] did in their immersion experiments—, if the data were normally distributed. All statistical analysis was performed using Python 3.8.3 with the packages Pandas 1.1.4 and SciPy 1.5.4. All visualizations were created using Seaborn 0.11.0. Additionally, we report the results of the open questions.

We are aware that neither the results of our statistical analysis nor those of our qualitative research should be used to infer knowledge about either the effect of real-time light estimation on immersion or energy consumption due to the small sample size. Rather, the analysis can give some indication as to the results of a full-scale study as well as enable us to evaluate the chosen methods and tools.

3   Results

The immersion scores per question according to the IEQ were calculated as ranging from a 1 for strongly disagree to a 5 for strongly agree. Scores for positively formulated questions were added, while scores for negatively formulated questions were subtracted to calculate the overall immersion score per participant and condition. Overall CPU usage at a given time was calculated by adding the CPU usage percentages for all activity kinds with the same timestamp. Since the System CPU Percentage Engineering Type [16] has a range from 0% to 100% multiplied by the number of CPU cores in the given device, this can result in values exceeding 100% (or 1 as in the formatting of our analysis).

Contrary to our first hypothesis **H1**, overall immersion scores did not differ significantly between both conditions (Mann-Whitney $U = 18.5$, $p = 0.241$) with the mean overall immersion score for the real-time light estimation *disabled* condition of 76.0 (SD = 6.6) even slightly exceeding the *enabled* condition with 74.7 (SD = 7.3) (see Figure 2).
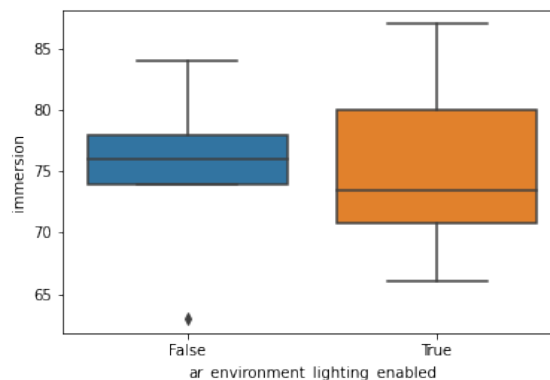

Figure 2: Immersion scores for the disabled and enabled conditions in our small pre-study ($n = 7$).

Regarding hypothesis **H1a**, there was no indication whether most participants in a full-scale study would notice a difference between both conditions as three of the seven participants in our pre-study stated they had not noticed any distinction. Within the group of participants that had answered to have noticed a difference, the

immersion scores between real-time light estimation enabled and disabled did not differ significantly (Mann-Whitney $U = 4.0$, $p = 0.155$) with mean immersion scores of 76.3 (SD = 4.9) for the enabled and 76.0 (SD = 11.4) for the disabled condition respectively.

Finally, contrary to our hypothesis **H2**, mean CPU usage did not significantly differ between both conditions (Mann-Whitney $U = 16.0$, $p = 0.405$) with 61.8% (SD = 6.7%) mean overall CPU usage in the disabled condition and 63.1% (SD = 6.6%) in the enabled condition. However, as can be seen in Figure 3, the distributions of overall CPU usage per condition across all experiments (i.e. without calculating a mean per experiment and condition first) did appear different. The enabled condition resulted in a significantly higher overall CPU usage even in our small pre-study (Mann-Whitney $U = 53272.5$, $p = 0.029$).
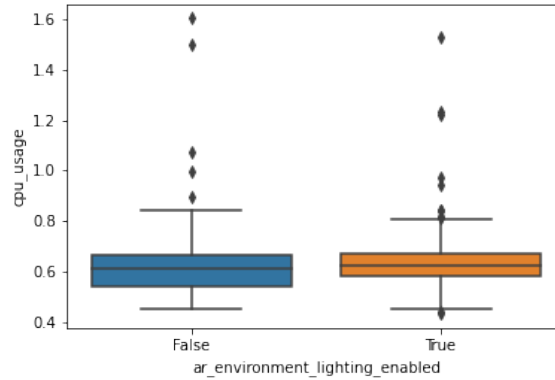

Figure 3: Overall CPU usage for the disabled and enabled conditions. The overall CPU usage can exceed 1 for devices with more than one CPU core (range from 0 to number of cores) as specified in the corresponding data type definition of the logging software [16].

4   Discussion

The results of our pre-study seem to indicate that there might not be a trade-off between improving immersion and increasing energy consumption when deploying real-time light estimation. Instead, it appears that immersion is largely unaffected by the feature while energy consumption as measured through CPU usage in our experiment might be slightly increased. This would lead us to conclude that real-time light estimation should not be employed by designers and developers, so that users' battery can be preserved or computational resources can be freed for alternative uses. We need, however, to look at the results in more detail to justify this conclusion.

As our results regarding hypothesis **H1** show, the immersion scores for our seven participants did not significantly differ between enabled and disabled real-time light estimation. We instead observed a quite wide distribution for both conditions ($SD_{enabled} = 7.3$, $SD_{disabled} = 6.7$). This could indicate that the difference between both lighting conditions is hardly perceivable when not put side by side and that other factors such as changes in the environment (e.g. noises or a change in natural lighting) diminish real-time light estimation's effect on immersion. This is further supported by our results regarding hypothesis **H1a**.

According to the questionnaire data as well as the comments from our study participants, the visual characteristics of the scene did not appear noticeably different between both conditions. Even when the setup of the experiment was explained to the participants, some could not recall any difference in appearance between the conditions. Only when toggling the corresponding render setting for a live scene resulting in an immediate change, the participants were able to recognize that the virtual objects were lit differently. Similarly, the immersion scores for both groups—those participants who claimed to have noticed a difference between both conditions and those who had not—did not differ significantly. It seems that the hypothesized improvement in scene realism brought about by real-time light estimation was not strong enough in our chosen application to be consciously perceived or to result in increased immersion as measured through the IEQ.

A similar impression emerges when examining CPU usage: Taking only the overall mean utilization into consideration, we would have to reject our hypothesis **H2**. Mean CPU usage appears largely unaffected by enabling or disabling real-time light estimation with the results of our Mann-Whitney U-test showing no

significant difference between both conditions and even the range of the two measurements being remarkably similar (disabled: 53.2% to 70.1%; enabled: 51.7% to 71.3%). When considering the distribution of all CPU usage data between both conditions, however (i.e. treating the overall CPU usage at any point of time during all experiments the same without aggregating the data for an experiment first), a slight but significant increase during enabled real-time light estimation can be observed. While this would generally support our hypothesis, the difference in CPU usage is so small that it would likely not result in a perceivable difference in energy consumption or battery life for users of the sample application. This conclusion would, however, need to be validated by another study.

This leaves us with the overall result of our pre-study: While the difference between both conditions can be noticed in a side-by-side comparison, enabling or disabling real-time light estimation does not have a significant effect on the appearance of AR scenes and consequently on immersion. And, while it statistically could be argued—even with our small sample size—that CPU usage and thereby energy consumption is affected by this choice, it would likely not result in a meaningful difference for the user experience of an AR application. It is, therefore, within designers' and developers' volition whether to enable or disable real-time light estimation for their applications according to their preferences or other decision factors than those examined in this study.

We are, however, aware of the numerous limitations of our pre-study and want to give as complete an account as possible of these weaknesses to prepare for a possible full-scale study with a greater number of participants investigating the same relationships.

5   Limitations and Future Work

The most pressing and most obvious limitation of our pre-study is the small sample size. We were largely limited to descriptive statistics and tests with low statistical power due to the fact that we could only gather data from seven participants. To reach higher levels of confidence, especially regarding any possible differences in immersion between both tested conditions, more participants would need to be recruited. This would likely also address another major issue regarding our sample: the lack of diversity. Due to our reliance on convenience sampling during a time where contact outside of one's immediate group of acquaintances is rather difficult or even prohibited, we were unable to recruit any female participants or participants outside the age range of students. For a full-scale study, more emphasis should be put on gathering a sample representing a larger part of the potential user group of AR applications.

The results might additionally be affected by the study design. Some participants noted that asking about the overall experience and any differences between both runs of the experiment would lead to biased answers since it provoked answers even if the participant would not have experienced significant differences. One participant suggested a third experiment variation in addition to the two randomized conditions tested in this study where the active render setting is not changed between both runs. In addition to these weaknesses of our chosen method, we observed some weaknesses of our tooling as well.

While conducting the pre-study, we experienced some difficulties with our chosen software. The modified sample application as well as the monitoring and logging software chosen had reliability issues that hindered the data gathering process. In one of the experiments, logging of the CPU usage suddenly stopped after circa 1.5 min. While other data such as network events that were logged automatically by the software could still be accessed, CPU data was non-retrievable. In another experiment, the logging software suddenly exited after the tester requested to stop and save the recording. Here again, CPU data was non-retrievable. The sample application, similarly, experienced some issues that seemed to relate to the Xcode IDE. In one experiment, the application failed to exit through the software and had to be manually shut down. Regarding less severe concerns, the sample application appeared to lack in terms of its feedback mechanisms. During multiple experiments, the probing phase (i.e. the phase during which the application gathers information about the environment, specifically to find a horizontal plane to use as an anchor for the virtual objects of the scene) which was indicated by an overlay with instructions to slowly move the device was exited without placing the virtual objects in an appropriate time frame. No obvious cause or pattern to this behavior could be identified. This lead to multiple restarts of the application which delayed several runs of the experiment.

During the experiments, several participants further experienced difficulties when interacting with the sample application. Since the application does support scaling and positioning of the virtual objects during the setup phase for increased usability but does not save these settings, the virtual cards were displayed in the original scale and position after each successful game resulting in visible irritation by the participants.

Additionally, the participants seemed to experience some discomfort holding the test device as multiple participants frequently adjusted their grip during the second run of the experiment. One participant even rested the device on a table to avoid having to hold it. It was, furthermore, noticeable that all participants moved the device during the trial phase of the experiment and some at the beginning of the first run, but all stopped moving the device eventually. They instead remained in a fixed position once a comfortable posture was discovered and only interacted with the application by tapping the screen. One participant tried multiple different postures, positions, distances, and orientations during the trial phase but, just as the other participants, stopped this exploratory behavior completely once the first run started. In an AR application, it could be argued, users should move to immerse themselves in the experience provided which our sample application failed to motivate. Instead, the lack of movement that most participants exhibited during the experiment might have lead to generally too low immersion levels to find a difference between both tested conditions. All of these observations seem to indicate generally questionable usability of the chosen application that should be addressed for a full-scale study.

Our discussions with the participants revealed that the reliability and usability issues we observed matched the participants' experiences. Multiple participants noted that they were irritated by the inconsistent scale and position of the virtual objects. Two participants noted that the application's simplicity might limit the potential immersion and that the lack of more complex objects and reflections might result in generally smaller differences in immersion between both runs. They, furthermore, suspected that this visual simplicity would make it harder for participants to notice any differences between both runs. Three participants stated unprompted that the device was uncomfortable to hold for the duration of the experiment due to its size and sharp edges.

Finally, multiple participants were briefly distracted during one or both runs of the experiment by noises or other people in the test environment. One participant attempted to start a conversation with the testers during one of the runs which was denied but still interrupted the run. This indicates that the study design was unable to prevent inconsistencies in how the experiment was run that might have affected the results. Since we did not have access to one consistent room with minimal distractions for our testing, we conducted the pre-study in multiple environments with various lighting conditions and possible disturbances. This was especially problematic when multiple participants were in the same room or when a common room was used. For a full-scale study, a consistent laboratory room with controlled lighting and minimal distractions should be used instead to avoid these inconsistencies and disturbances.

In regard to the procedure, the lack of a clear script or protocol to explain the experiment in detail and consistently to each participant became clear when multiple participants stopped after their first completed round of the sample game during the first tested condition. Despite the explanations given by the testers, they were unaware of their goal to complete as many rounds of the game with as few errors as possible within the given time frame. This shows that the explanations were unable to provide participants with sufficiently understandable descriptions of the tasks that were expected from them. For a full-scale study, a clear protocol to be read aloud by the tester should be considered. Moreover, as one participant pointed out, the inclusion of a placebo condition might be an appropriate addition to further control bias. While the randomization is well suited to reduce the effect of order on the overall immersion scores, introducing a condition where participants are shown the same lighting in both conditions could reveal further information about the study design's effect on the results.

Regarding the questionnaire that accompanied the experiment, there are four observations that we want to highlight here. First, we noticed that multiple participants were slightly irritated by the fact that we used one Google Forms form for all information regarding the experiment (incl. the consent form and data to be filled by the tester). This interrupted the flow of the experiment and can easily be avoided by splitting the questionnaire into one to be filled by the participant and one to be filled by the tester which can be joined by an anonymous ID. Second, we observed that some participants might have been affected in their answers to the IEQ by the testers' presence in the room when they filled the questionnaire. Two participants, for example, asked for an explanation of one of the questions of the IEQ. In order to not affect the results of the questionnaire, the participants should answer the questions to the best of their knowledge and understanding. One simple way to mitigate this problem would therefore be to leave the room where the experiment is being conducted for the duration of the questionnaire. Third, we became aware that our definition of "noticing a difference between both tested conditions" using a very limited dictionary approach was too wide and therefore might have lead to some inconsistencies in understanding. This might have influenced the results of our analysis regarding hypothesis **H1a**, since we did not find any clear indication whether users would or would not notice the difference in

lighting. For a full-scale study, this should be further defined and included in the form to be answered by the participant. Finally, the choice of the IEQ might have to be re-considered since some of the questions (e.g. "To what extent was your sense of being in the game environment stronger than your sense of being the real world?") are less suited for an AR game. This might be one of the reasons why our preliminary results regarding hypothesis **H1** seem to indicate no significant effect of real-time light estimation on overall immersion. Since available alternatives exist that are in some cases even developed specifically for AR applications [8], these should be reviewed before conducting a full-scale study based on the IEQ.

Similarly, it can be argued that CPU usage is an unsuitable measure for energy consumption since we did not examine their assumed correlation for our tested application. This objection is further supported by the fact that, typically, the graphics processing unit (GPU) handles most computation for rendering and that overall CPU usage is not only affected by the sample application. Multiple background services might mask the effect of real-time light estimation on the CPU. While the measure was chosen because of (1) its granular logging possibilities, (2) because of the CPU's assumed sensitivity to graphics rendering as the test device uses an integrated system on a chip [21], and while (3) it seems to be affected by the rendering condition in our preliminary testing (see the analysis regarding **H2** in Section 3), these concerns should be addressed by further research considering alternative measures and frameworks. Additionally, it could be tested whether CPU usage and energy consumption correlate using more direct but less granular measures like Xcode's Instrument's Abstract Power Measurement [22].

Finally, it should be reviewed whether the study design and tooling provided by this project are economical enough for a full-scale study. Since the setup requirements are quite high—a consistently available room with minimal distractions and controlled lighting, a consistent and high-powered test device, at least one tester per participant, and both a questionnaire as well as an experiment—and since the results might be fast outdated due to fast-changing APIs and their underlying technology, a full-scale study might consider splitting off or re-designing parts of the study, for example by testing the effect of real-time light estimation on energy consumption in a separate, more controlled study without participants. If the study design is deemed economical, it should be considered whether to expand the scope by including multiple devices, multiple framework APIs as well as multiple API versions to test for their effect on the results.

6   Conclusion

While our pre-study has several limitations in regards to general knowledge that can be inferred from our results, it seems to give some valuable indications for our research questions. Concerning research question **RQ1** whether users of AR applications would experience a difference in immersion between enabled and disabled real-time light estimation, the results of our analysis seem to suggest that there might not be a considerable effect of real-time light estimation on immersion. As both the answers to the IEQ and the comments of our participants indicate, the difference in colors and contrast caused by enabling real-time light estimation might generally not be strong enough to be noticeable without directing users' attention to it. Regarding research question **RQ1a** whether participants would be able to tell that lighting was the cause for any perceived difference between both conditions, this means that—while participants might in fact perceive a difference between both conditions—-the difference in lighting is so subtle that users have difficulty articulating what caused this impression. Finally, in regards to research question **RQ2** how real-time light estimation would affect CPU utilization, our results point to a slight increase when enabling real-time light estimation. This effect might, however, be so small as to be unnoticeable in real-world usage. For a more meaningful account, additional measures that more closely relate to device energy consumption and battery life should be monitored as well. Overall, the proposed method and tools are suitable to conduct a full-scale study on the effects of real-time light estimation on immersion and energy consumption that builds on the results of this pre-study, but each of the parts should carefully be considered, evaluated, and possibly modified before investing heavily in such an endeavor. This study should be used as a starting point for designing such a full-scale study.

References

[1] Apple, "RealityKit - Augmented Reality," 2020. [Online]. Available: https://developer.apple.com/augmented-reality/realitykit/ [Accessed: 2020-11-23].

[2] ——, "ARLightEstimate | Apple Developer Documentation," 2020. [Online]. Available: https://developer.apple.com/documentation/arkit/arlightestimate [Accessed: 2020-09-03].

[3] ——, "disableAREnvironmentLighting | Apple Developer Documentation," 2020. [Online]. Available: https://developer.apple.com/documentation/realitykit/arview/renderoptions/3368184-disablearenvironmentlighting [Accessed: 2020-11-23].

[4] J. Steed, A. Krauss, M. Scavezze, W. Zhang, A. Tomlin, T. Ambrus, B. Mount, S. Latta, and R. Hastings, "Constructing augmented reality environment with pre-computed lighting," US Patent US9 524 585B2, Dec., 2016. [Online]. Available: https://patents.google.com/patent/US9524585B2/en

[5] J. L. Gabbard, J. E. Swan, and D. Hix, "The Effects of Text Drawing Styles, Background Textures, and Natural Lighting on Text Legibility in Outdoor Augmented Reality," *Presence: Teleoperators and Virtual Environments*, vol. 15, no. 1, pp. 16–32, Feb. 2006. doi: 10.1162/pres.2006.15.1.16 Publisher: MIT Press. [Online]. Available: https://doi.org/10.1162/pres.2006.15.1.16

[6] M. Aittala, "Inverse lighting and photorealistic rendering for augmented reality," *The Visual Computer*, vol. 26, no. 6, pp. 669–678, Jun. 2010. doi: 10.1007/s00371-010-0501-7. [Online]. Available: https://doi.org/10.1007/s00371-010-0501-7

[7] D. Johnson, M. J. Gardner, and R. Perry, "Validation of two game experience scales: The Player Experience of Need Satisfaction (PENS) and Game Experience Questionnaire (GEQ)," *International Journal of Human-Computer Studies*, vol. 118, pp. 38–46, Oct. 2018. doi: 10.1016/j.ijhcs.2018.05.003. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1071581918302337

[8] Y. Georgiou and E. A. Kyza, "The development and validation of the ARI questionnaire: An instrument for measuring immersion in location-based augmented reality settings," *International Journal of Human-Computer Studies*, vol. 98, pp. 24–37, Feb. 2017. doi: 10.1016/j.ijhcs.2016.09.014. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1071581916301306

[9] M. Csikszentmihalyi, *Flow: the psychology of optimal experience*, 1st ed. New York: Harper & Row, 1990. ISBN 978-0-06-016253-5

[10] C. Jennett, A. L. Cox, P. Cairns, S. Dhoparee, A. Epps, T. Tijs, and A. Walton, "Measuring and defining the experience of immersion in games," *International Journal of Human-Computer Studies*, vol. 66, no. 9, pp. 641–661, Sep. 2008. doi: 10.1016/j.ijhcs.2008.04.004. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1071581908000499

[11] A. Dey, M. Billinghurst, R. W. Lindeman, and J. E. I. Swan, "A Systematic Review of 10 Years of Augmented Reality Usability Studies: 2005 to 2014," *Frontiers in Robotics and AI*, vol. 5, 2018. doi: 10.3389/frobt.2018.00037 Publisher: Frontiers. [Online]. Available: https://www.frontiersin.org/articles/10.3389/frobt.2018.00037/full

[12] A. Dünser, R. Grasset, and M. Billinghurst, "A survey of evaluation techniques used in augmented reality studies," in *ACM SIGGRAPH ASIA 2008 courses on - SIGGRAPH Asia '08*. Singapore: ACM Press, 2008. doi: 10.1145/1508044.1508049 pp. 1–27. [Online]. Available: http://portal.acm.org/citation.cfm?doid=1508044.1508049

[13] B. G. Witmer and M. J. Singer, "Measuring Presence in Virtual Environments: A Presence Questionnaire," *Presence: Teleoperators and Virtual Environments*, vol. 7, no. 3, pp. 225–240, Jun. 1998. doi: 10.1162/105474698565686 Publisher: MIT Press. [Online]. Available: https://doi.org/10.1162/105474698565686

[14] M. Cobb, "maxxfrazer/RealityKit-CardFlip," Sep. 2020, original-date: 2019-07-17T18:39:03Z. [Online]. Available: https://github.com/maxxfrazer/RealityKit-CardFlip [Accessed: 2020-09-15].

[15] Apple, "Xcode - Features," 2020. [Online]. Available: https://developer.apple.com/xcode/features/ [Accessed: 2020-09-15].

[16] ——, "System CPU Percentage Engineering Type - Instruments Developer Help," 2020. [Online]. Available: https://help.apple.com/instruments/developer/mac/current/#/dev167969024 [Accessed: 2020-09-12].

[17] ——, "Energy Impact Engineering Type - Instruments Developer Help," 2020. [Online]. Available: https://help.apple.com/instruments/developer/mac/current/#/dev63504527 [Accessed: 2020-09-12].

[18] M. van den Haak, M. De Jong, and P. Jan Schellens, "Retrospective vs. concurrent think-aloud protocols: Testing the usability of an online library catalogue," *Behaviour & Information Technology*, vol. 22, no. 5, pp. 339–351, Sep. 2003. doi: 10.1080/0044929031000. [Online]. Available: http://www.tandfonline.com/doi/abs/10.1080/0044929031000

[19] E. A. Boyle, T. M. Connolly, T. Hainey, and J. M. Boyle, "Engagement in digital entertainment games: A systematic review," *Computers in Human Behavior*, vol. 28, no. 3, pp. 771–780, May 2012. doi: 10.1016/j.chb.2011.11.020. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0747563211002640

[20] A. I. Nordin, J. Ali, A. Animashaun, J. Asch, J. Adams, and P. Cairns, "Attention, time perception and immersion in games," in *CHI '13 Extended Abstracts on Human Factors in Computing Systems on - CHI EA '13*. Paris, France: ACM Press, 2013. doi: 10.1145/2468356.2468551. ISBN 978-1-4503-1952-2 p. 1089. [Online]. Available: http://dl.acm.org/citation.cfm?doid=2468356.2468551

[21] Apple, "Apple introduces 10.5-inch iPad Pro, updates 12.9-inch model," 2017. [Online]. Available: https://www.apple.com/li/newsroom/2017/06/ipad-pro-10-5-and-12-9-inch-models-introduces-worlds-most-advanced-display-breakthrough-performance/ [Accessed: 2021-01-02].

[22] ——, "Abstract Power Measurement Engineering Type - Instruments Developer Help," 2020. [Online]. Available: https://help.apple.com/instruments/developer/mac/current/#/dev68944042 [Accessed: 2020-11-29].

A Appendix: Participants' Consent Form

Informed Consent Form

The following is an informed consent document for the purpose of freeing students/faculty of any liability while performing a research study with human participants. The study, performed by Stefano Formicola and Christoph Albert Johns, aims at illuminating the relationship between immersion and energy impact in deploying advanced augmented reality (AR) features, thus empowering developers and designers to decide the most appropriate feature sets for their applications considering both user experience and sustainability aspects.

The experiment will consist of two three-minute session of an Augmented Reality game, with and without a specific setting enabled, and data collection and measurement. Following, you will find a more precise description of collected data.

All information you provide will remain confidential and will not be associated with your name. If for any reason during this study you do not feel comfortable, you may leave the experiment at any time and your information will be discarded. Your participation in this study will require approximately 30 minutes. When this study is complete you will be provided with the results of the experiment if you request them, and you will be free to ask any questions.

If you have any further questions concerning this study please feel free to contact us through email: Stefano Formicola at formico@kth.se or Christoph Albert Johns at cajohns@kth.se. Your participation is solicited, yet strictly voluntary. All information will be kept confidential and your name will not be associated with any research findings and is only recorded to remove your data if so required. You can withdraw your consent at any time.

1. Subject
   Name and Surname of the subject involved in the experiment. This will not be shared nor will the data be associated with your name.

2. Date
   DD/MM/YYYY

3. I agree to share: (checkbox)
   – Device measurements (i.e. CPU usage, overall duration) during my usage
   – My questionnaire answers

4. Research study introduction (checkbox)
   – I agree the experiment has been fully explained to me, as well as the data collected and my right to withdraw in any moment.

5. Research study consent (checkbox)
   – I have read all the information of this form and am willing to participate in the research study.

B   Appendix: Immersive Experience Questionnaire

Please answer the following questions by selecting the relevant number. In particular, remember that these questions are asking you about how you felt at the end of the game.

1. To what extent did the game hold your attention?
   Not at all   1   2   3   4   5   A lot

2. To what extent did you feel you were focused on the game?
   Not at all   1   2   3   4   5   A lot

3. How much effort did you put into playing the game?
   Very little   1   2   3   4   5   A lot

4. Did you feel that you were trying you best?
   Not at all   1   2   3   4   5   Very much so

5. To what extent did you lose track of time?
   Not at all   1   2   3   4   5   A lot

6. To what extent did you feel consciously aware of being in the real world whilst playing?
   Not at all   1   2   3   4   5   Very much so

7. To what extent did you forget about your everyday concerns?
   Not at all   1   2   3   4   5   A lot

8. To what extent were you aware of yourself in your surroundings?
   Not at all   1   2   3   4   5   Very aware

9. To what extent did you notice events taking place around you?
   Not at all   1   2   3   4   5   A lot

10. Did you feel the urge at any point to stop playing and see what was happening around you?
    Not at all   1   2   3   4   5   Very much so

11. To what extent did you feel that you were interacting with the game environment?
    Not at all   1   2   3   4   5   Very much so

12. To what extent did you feel as though you were separated from your real-world environment?
    Not at all   1   2   3   4   5   Very much so

13. To what extent did you feel that the game was something you were experiencing, rather than something you were just doing?
    Not at all   1   2   3   4   5   Very much so

14. To what extent was your sense of being in the game environment stronger than your sense of being in the real world?
Not at all   1   2   3   4   5   Very much so

15. At any point did you find yourself become so involved that you were unaware you were even using controls?
Not at all   1   2   3   4   5   Very much so

16. To what extent did you feel as though you were moving through the game according to you own will?
Not at all   1   2   3   4   5   Very much so

17. To what extent did you find the game challenging?
Not at all   1   2   3   4   5   Very difficult

18. Were there any times during the game in which you just wanted to give up?
Not at all   1   2   3   4   5   A lot

19. To what extent did you feel motivated while playing?
Not at all   1   2   3   4   5   A lot

20. To what extent did you find the game easy?
Not at all   1   2   3   4   5   Very much so

21. To what extent did you feel like you were making progress towards the end of the game?
Not at all   1   2   3   4   5   A lot

22. How well do you think you performed in the game?
Very poor   1   2   3   4   5   Very well

23. To what extent did you feel emotionally attached to the game?
Not at all   1   2   3   4   5   Very much so

24. To what extent were you interested in seeing how the game's events would progress?
Not at all   1   2   3   4   5   A lot

25. How much did you want to "win" the game?
Not at all   1   2   3   4   5   Very much so

26. Were you in suspense about whether or not you would win or lose the game?
Not at all   1   2   3   4   5   Very much so

27. At any point did you find yourself become so involved that you wanted to speak to the game directly?
Not at all   1   2   3   4   5   Very much so

28. To what extent did you enjoy the graphics and the imagery?
Not at all   1   2   3   4   5   A lot

29. How much would you say you enjoyed playing the game?
Not at all   1   2   3   4   5   A lot

30. When interrupted, were you disappointed that the game was over?
Not at all   1   2   3   4   5   Very much so

31. Would you like to play the game again?
Definitely not   1   2   3   4   5   Definitely yes